

制药设备故障大数据分析 with 预测技术研究

王松

(北京协和药厂有限公司 北京 102600)

摘要: 随着制药行业的发展, 制药设备的稳定运行对于保证药品生产的连续性和产品质量至关重要。本文重点探讨了制药设备常见的故障类型, 详细描述了故障数据的收集和处理方法, 并利用大数据技术对这些数据进行深入分析, 从中提取有价值的信息。进一步地, 通过采用机器学习和深度学习等先进方法, 研究了故障预测技术, 并对不同的预测模型进行了比较和选择。通过实证研究, 验证了所提出方法的准确性和实用性, 为制药设备的稳定运行和维护提供了有力的技术支持。

关键词: 制药设备; 故障预测; 大数据分析

0 引言

随着现代科技的迅速发展, 尤其是大数据技术的普及应用, 为各行各业带来了前所未有的机遇和挑战。在制药行业中, 药品生产设备作为关键的生产要素, 其稳定、高效地运行对于确保生产过程的流畅性和药品质量起到了至关重要的作用。然而, 随着生产规模的扩大和设备复杂度的增加, 设备故障已成为制约生产效率和产品质量的一个重要因素。如何及时、准确地预测和防范这些故障, 成为业界亟待解决的问题。

1 制药设备常见故障类型与数据收集

1.1 制药设备常见的故障类型及其产生的原因

制药设备在生产过程中, 由于其高度复杂的机械构造和对精准操作的需求, 容易出现多种故障。常见的故障类型包括机械部件磨损、电气控制系统故障、传感器失效、密封件泄漏、加热和冷却系统不稳定等。这些故障的产生原因可以归纳为:

(1) 设备长时间运行带来的磨损, 如轴承、齿轮和其他活动部件的磨损。

(2) 不恰当的操作和维护。例如, 操作人员的非标准操作或忽视日常维护。

(3) 材料老化。例如, 密封圈和橡胶件在长时间使用后出现硬化或断裂。

(4) 环境因素。例如, 温度、湿度的变化导致电子元件失效或腐蚀。

(5) 制药原料中的杂质或化学反应产物可能导致某些设备部件堵塞或腐蚀^[1]。

对于这些故障, 及时的预测和维护是确保生产流程稳定和提高产品质量的关键。

1.2 数据收集方式、数据源、数据的初步处理

制药设备的数据收集是故障预测的基石, 如图1所示, 主要包括以下几方面内容。

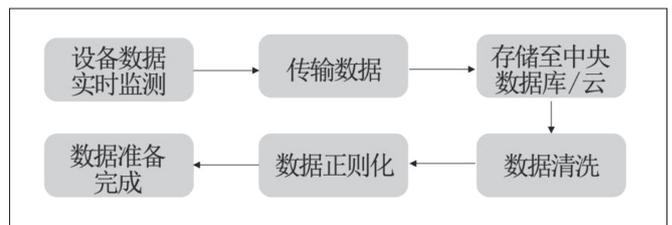


图1 数据初步处理流程

首先, 数据的收集方式主要基于高度灵敏的传感器, 如热电偶、压力传感器和流速计等, 其输出通常是电压或电流, 与设备的实际运行参数(例如, T 代表温度, P 代表压力)存在线性关系: $T=aV+b$ 。这些数据经由数据采集卡实时读取, 再传输到中央数据库或云端储存。其次, 数据源不仅来源于设备的实时监测, 还包括生产日志、维修记录及历史故障数据库。一旦数据被集中存储, 初步处理便开始执行, 包括数据清洗和正则化。

数据清洗的主要任务是去除异常值, 通过 IQR 方法找到并剔除落在 $Q1-1.5 \times IQR$ 或 $Q3+1.5 \times IQR$ 之外的值。其中, $Q1$ 为下四分位数, $Q3$ 为上四分位数, IQR 为四分位距。

正则化则是确保所有数据都在同一个尺度上，主要采用最小-最大缩放法，其公式如式(1)所示：

$$X_{nor} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

式中： X_{nor} - 缩放后的特征值；
 X - 原始特征值；
 X_{min} - 特征值的最小值；
 X_{max} - 特征值的最大值。

经过这些初步处理后，数据为后续的分析 and 建模做好了准备，确保准确性和可靠性^[2]。

2 大数据技术在故障分析中的应用

2.1 数据清洗与预处理的重要性

在制药设备故障分析的大数据技术应用中，数据清洗与预处理被视为关键性的前置步骤。原始的故障数据往往是由多种传感器、日志文件和操作员报告等多源途径产生的，这使得数据中常常充斥着噪声、异常值、重复数据和格式不一致性。这些问题不仅可能导致故障诊断的准确性受损，还会增加计算的复杂性和开销。特别是在应用统计模型和机器学习算法进行故障预测时，未经清洗和预处理的数据可能会导致模型的误导，从而影响决策效果。因此，首先进行数据清洗，以去除错误测量、消除重复值、填补缺失数据和纠正数据不一致性，是确保高质量分析的基石。接着，在预处理阶段，经常采用标准化或归一化技术来调整数据尺度，确保数据在统一的量纲上。这样做可以优化模型的性能，加速算法的收敛，为后续的深入分析和建模奠定坚实基础。总之，只有通过精细的数据清洗与预处理，才能确保制药设备故障分析的准确性、高效性和实用性，为企业带来真正的决策价值。

2.2 数据分析的方法：描述性分析、关联性分析等

在制药设备故障分析中，采用多种数据分析方法来对大数据进行深度挖掘，其中描述性分析和关联性分析是最基础和常用的两种方法。

描述性分析主要是通过计算各种统计量来描述数据的基本特性。例如，对于一个故障数据集，可能关心故障发生的平均频率等。以故障发生的平均频率为例，常用的统计量包括平均数、方差、标准差、中位数和众数等。

平均数：

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

式中： n - 故障频率次数；
 μ - 故障频率次数的平均次数；
 x_i - 第 i 次故障特征。
方差：

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (3)$$

标准差：

$$\sigma = \sqrt{\sigma^2} \quad (4)$$

这些统计量提供了数据的“大致轮廓”，有助于快速了解数据集的基本情况^[3]。

关联性分析则更进一步，它旨在研究两个或多个变量之间的关系。例如，想要了解某种特定故障与设备的工作时长是否有关。这时，可以采用皮尔逊相关系数来度量两个连续变量之间的线性关系，其计算公式为：

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

其中， x_i 和 y_i 分别是两个变量的观测值，而 \bar{x} 和 \bar{y} 是它们的均值。当 $r=1$ 时表示完全正相关， $r=-1$ 时表示完全负相关， $r=0$ 时表示无关。

此外，对于类别数据，可以使用卡方检验来判断两个类别变量是否相关。通过这些关联性分析方法，可以更深入地挖掘数据背后的潜在规律，为故障预测和决策提供科学依据。

2.3 从大量故障数据中提取有用的模式和趋势

在制药设备故障数据的深度分析中，识别出有用的模式与趋势是至关重要的。通过时间序列分析，能够明确故障的周期性特征，如傅里叶变换揭示的特定时间段内的故障频率，从而优化维护策略。同时，利用聚类算法，如 K-means，可以将具有相似特性的故障数据有效地分组，这有助于更深入地了解常见的故障模式，如与特定环境条件或操作相关的故障。此外，关联规则挖掘，例如，利用 Apriori 算法，能够发现不同故障之间的潜在关系，为故障诊断提供线索。趋势的预测，如通过回归分析，不仅能够洞

察未来的可能故障，还为资源分配和备件采购提供参考。而异常检测技术，如孤立森林，为早期识别出与常态偏离的故障提供了有力工具。

3 故障预测技术及方法

3.1 机器学习、深度学习等预测方法

机器学习和深度学习示意图如图2所示。

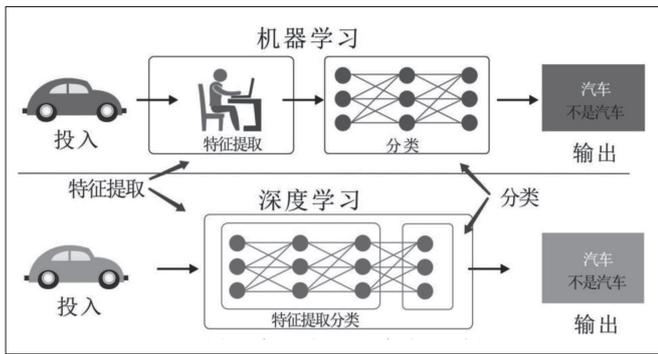


图2 机器学习和深度学习示意图

故障预测是工业维护的核心领域，而机器学习和深度学习在此方面展现出了卓越的性能。机器学习，如支持向量机（SVM）、随机森林和梯度提升树（GBT），主要依赖于历史数据和特征工程，通过学习历史数据中的模式来预测未来的故障。这些算法对特征的选择和调优尤为敏感，需要大量的领域知识。而深度学习，特别是卷积神经网络（CNN）和循环神经网络（RNN），则可以自动从数据中提取有意义的特征，使得预测更为准确。LSTM，一种特定的RNN结构，由于其在处理时间序列数据方面的优势，已广泛应用于设备的故障预测中。同时，自编码器深度学习模型在未标记数据的异常检测中也展现出强大的实力^[4]。总的来说，机器学习为故障预测提供了快速而准确的方法，而深度学习则能够处理更为复杂的数据结构和模式，为故障预测提供了更深层次的见解。这两者结合，为制药设备的故障预测提供了坚实的技术基础。

3.2 特征工程在故障预测中的作用

特征工程在故障预测中起到了至关重要的作用。它是将原始数据转化为机器学习算法可以利用的特征的过程，是连接领域知识与数据分析的桥梁。在故障预测的上下文中，特征工程的主要任务是从复杂的机器或设备操作数据中提取有关其健康状况和潜在故障的有意义信息。

首先，原始数据往往包含大量的噪声和冗余信息，

特征工程可以通过各种统计方法和算法来消除这些不必要的信息，仅保留与预测任务最相关的特征。例如，通过滑动窗口技术计算设备传感器数据的移动平均值，可以有效地去除短期的随机波动，凸显出设备的长期趋势。

其次，领域知识在特征工程中尤为重要。对于特定的设备和其潜在故障，领域专家可以识别出最关键的参数和指标。例如，在某些机械设备中，温度、压力或振动的突然变化可能预示着即将发生的故障。通过对这些关键参数进行特征工程，可以提高故障预测的准确性和鲁棒性。

3.3 不同预测模型的比较与选择

在故障预测领域，选择适当的预测模型是确保准确性和效率的关键。线性回归因其对线性关系的高效处理而受到青睐，但面对非线性复杂数据时可能无法处理。与此相对，决策树和其集成版本——随机森林，能够递归地分割数据，更好地捕捉复杂模式。支持向量机在高维空间中寻找最优分割平面，但计算速度可能受限于大规模数据。深度学习，特别是卷积神经网络和循环神经网络，已证明在处理含有难以识别模式的大型数据集时表现卓越^[5]。当预测的精确性和模型的可解释性成为首要考虑时，集成学习方法，如XGBoost等提升算法，将成为首选。最终选择模型时，还需考虑数据规模、特征复杂性、计算资源等因素，并利用交叉验证等技术评估模型性能，以确保故障预测的可靠性和鲁棒性。

4 实证研究

4.1 选取具体的制药设备进行故障数据分析

在本次研究中，选择了一台大型的制药混合机，该混合器在制药生产线中扮演关键角色，主要用于混合多种原料制成药物。此混合器自投入使用以来，运行时间已超过10年。近期，操作人员在使用过程中发现，当混合器运行一段时间后，会出现轻微的异响，且在连续工作数小时后，机器温度明显升高。同时，混合后的药物粉末有时会出现不均匀的情况，这影响了生产质量和效率。此外，混合器的开关机时间变得不稳定，有时在没有预警的情况下自动停机。为了保证制药生产的稳定和药物的质量，对该混合器的故障数据进行了深入的分析。

4.2 采用前述方法进行故障预测

对于该药物混合机的故障数据，首先采用了数据

清洗与预处理的步骤，剔除了其中的异常值和无关信息，将数据标准化以消除数据的量纲影响。在这个基础上，使用描述性分析得到了该设备在各种工作状态下的平均工作参数，如平均温度、平均振动频率等，并通过关联性分析探索了这些参数之间的内在联系。

进一步地，基于机器学习的方法，如支持向量机、随机森林 (Random Forest) 及深度学习的方法，如卷积神经网络，构建了故障预测模型。为了确保模型的预测准确性，进行了特征工程，挑选了与故障最为关联的特征，如混合器的启动次数、连续工作时间、温度升高的速率等。

在模型训练的过程中，使用交叉验证技术优化了模型参数，以获得最佳的预测效果。当模型构建完成后，使用前所未见的测试数据进行了模型评估，评估结果显示，深度学习模型在故障预测方面相较于传统的机器学习方法具有更高的准确性。

4.3 分析预测结果的准确性与实用性

预测结果分析见表 1。

表 1 预测结果分析

故障类型	实际发生次数 / 次	预测发生次数 / 次	预测准确度 / %
温度过高	10	12	83.3
振动异常	8	7	87.5
轴承磨损	5	6	83.3
电动机过载	3	4	75

从表 1 中可以观察到，对于制药混合器的故障预测，模型整体上表现得相对准确。对“温度过高”这一故障的预测准确度达到了 83.3%，尽管预测发生次数略高于实际发生次数，但其偏差在可接受的范围内。相似地，对“振动异常”和“轴承磨损”的预测准确度也达到了较高的水平，分别为 87.5% 和 83.3%。但值得注意

的是，“电动机过载”这一故障的预测准确度稍低，为 75%。这可能是因为某些微小的电动机过载现象并没有被成功捕捉，或是存在其他外部干扰导致的预测误差。总的来说，这些预测结果为制药生产提供了有力的决策支持，能够帮助企业提前采取措施，避免潜在的设备故障，从而确保生产流程的稳定和生产效率的提高。尽管存在一定的预测误差，但整体上，该模型具有很高的实用性和应用价值。

5 结语

随着制药工业的持续发展和技术的日趋复杂化，确保设备的稳定运行变得尤为关键。本研究为制药设备的故障预测和管理提供了一种新的、高效的方法论框架。随着技术的进一步发展，预期未来将有更多的算法和工具应用于此领域，为制药工业带来更大的经济和社会效益。

参考文献：

[1] 周元高，曾祥君．基于预测技术的设备机群故障分析与循环改进 [J]．科技和产业，2021，21(06)：284-287.
 [2] 任亚伟．基于关联规则和 BP 神经网络的飞机设备故障分析与预测 [D]．济南：山东师范大学，2020.
 [3] 谢国财，温锐，陈琛．基于模糊神经网络的高压电力设备故障预测模型 [J]．电网与清洁能源，2022，38(09)：120-125.
 [4] 赵军富，靳永胜，李建军．物联网下智能制造设备故障预测分析 [J]．内蒙古科技大学学报，2023，42(02)：123-127.
 [5] 巫小燕，刘永明，谢鹏，等．基于支持向量机的 CTD 烟丝干燥设备故障预测研究 [J]．菏泽学院学报，2022，44(05)：58-63.

作者简介：王松 (1988.04-)，男，汉族，河北保定人，本科，助理工程师，研究方向：制药设备维保维修。